

A Method for Estimating Cost Savings for Population Health Management Programs

Shannon M.E. Murphy, John McGready, Michael E. Griswold, and Martha L. Sylvia

Objective. To develop a quasi-experimental method for estimating Population Health Management (PHM) program savings that mitigates common sources of confounding, supports regular updates for continued program monitoring, and estimates model precision.

Data Sources. Administrative, program, and claims records from January 2005 through June 2009.

Data Collection/Extraction Methods. Data are aggregated by member and month.

Study Design. Study participants include chronically ill adult commercial health plan members. The intervention group consists of members currently enrolled in PHM, stratified by intensity level. Comparison groups include (1) members never enrolled, and (2) PHM participants not currently enrolled. Mixed model smoothing is employed to regress monthly medical costs on time (in months), a history of PHM enrollment, and monthly program enrollment by intensity level. Comparison group trends are used to estimate expected costs for intervention members. Savings are realized when PHM participants' costs are lower than expected.

Principal Findings. This method mitigates many of the limitations faced using traditional pre-post models for estimating PHM savings in an observational setting, supports replication for ongoing monitoring, and performs basic statistical inference.

Conclusion. This method provides payers with a confident basis for making investment decisions.

Key Words. Cost savings, cost effectiveness, mixed model smoothing, population health management, disease management

Rising health care costs are a major concern facing employers, managed care organizations, and other payers of health care services across the nation. Those who use the highest amount of health care resources are often those with multiple chronic conditions (Hoffman, Rice, and Sung 1996; Wolff,

Starfield, and Anderson 2002), and a large portion of their costs are thought to be preventable through behavior modification and symptom management (Wolff, Starfield, and Anderson 2002). In an effort to contain costs for multimorbid populations, Population Health Management (PHM) programs have flourished in the past decade (Wheatley 2002; Juster et al. 2009). PHM has been described as “a proactive, organized, and cost-effective approach to prevention that utilizes newer technologies to help reduce morbidity while improving health status, health service use, and personal productivity of individuals in defined populations” (Chapman and Pelletier 2004). For managing those with chronic illness, interventions typically consist of nurse case management, telemonitoring, health coaching, motivational interviewing, and condition-specific health education to inform individuals about their condition and to assist them in developing self-management skills. The goals of PHM include improved health and productivity, reduction in modifiable risk factors, promotion of appropriate health care utilization, and reduction of preventable hospitalizations, which should ultimately result in lower health care costs (Johnson 2003; Kindig and Stoddart 2003; Villagra and Ahmed 2004; Seto 2008; Gary et al. 2009; Hibbard, Greene, and Tusler 2009).

Despite the belief in the success of PHM programs, there is little evidence to suggest that they significantly reduce health care costs. The cost-effectiveness results in the literature tend to be inconclusive or mixed, with some studies demonstrating savings (Sidorov et al. 2002; Cousins and Liu 2003; Villagra and Ahmed 2004; Sylvia et al. 2008; Cloutier et al. 2009; Dall et al. 2010), whereas others do not (Galbreath et al. 2004; Turner et al. 2008; Nyman, Barleen, and Dowd 2009; Peikes et al. 2009). Results from these prior studies are often suspect, however, due to the frequent use of weak research designs. The validity issue is particularly concerning given that studies lacking in methodological rigor tend to report higher savings and return on investments (ROIs) than studies with more rigorous designs (Linden 2006; Maciejewski, Chen, and Au 2009). For example, studies that have demonstrated significant cost reductions are often those using weaker research methodologies such as pre-post designs without a comparison group, small sample sizes, inadequate observation lengths, and oversimplified statistical methods (Wagner 1998;

Address correspondence to Shannon M. E. Murphy, M.A., Johns Hopkins HealthCare LLC, 6704 Curtis Court, Glen Burnie, MD 21060; e-mail: smurphy@jhhc.com. John McGready, Ph.D., is with the Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. Michael E. Griswold, Ph.D., is with the Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, MS. Martha L. Sylvia, Ph.D., M.B.A., R.N., is with the School of Nursing, Johns Hopkins University, Baltimore, MD.

Selby et al. 2003; Congressional Budget Office 2004; Fitzner et al. 2004). On the other hand, studies using randomized techniques often failed to detect significant cost savings related to PHM (Galbreath et al. 2004; Turner et al. 2008; Peikes et al. 2009). Even in experimental settings, however, the use of an intention-to-treat approach that does not consider intensity or duration of participation in the delivered intervention may dilute the impact of PHM programs and hinder the ability to detect significant cost savings.

To confidently evaluate the fiscal impact of PHM programs, an improved method is needed. In health care settings, an experimental design is not often feasible due to ethical or contractual issues. The National Committee for Quality Assurance (NCQA) has recognized that randomized control trials are impractical within health plans (Fetterolf, Wennberg, and Devries 2004). Consequently, most researchers are limited to observational studies using some form of pre-post design (Wilson et al. 2004). The difficulty is that these types of designs are plagued by multiple confounding factors that need to be addressed to preserve the internal validity of the findings.

One group of pre-post designs, although the most prominent design used for program evaluation (Johnson 2003; Linden, Adams, and Roberts 2003), can be limited by a number of confounders, including history, maturation, and regression to the mean. *History* effects may call into question the results if an event occurred between the pre and post measurements that impacted the outcome but was unrelated to the intervention or not of specific interest (Cook and Campbell 1979). *Maturation* involves the natural changes within participants through the passage of time, such as aging, tiredness, and hunger. This threat is of particular concern for PHM program evaluation given that a participant may improve naturally over time, regardless of whether they participate in the program (Cook and Campbell 1979). *Regression to the mean* suggests that extreme observations (in either direction) will eventually move toward the mean (Galton 1886), particularly if the observations are extreme due to measurement error (Cook and Campbell 1979). When PHM participants are selected based on extreme characteristics, such as a catastrophic event that leads to an expensive hospitalization, it is likely that participant costs will decrease over time regardless of the intervention (Cook and Campbell 1979; Fetterolf, Wennberg, and Devries 2004; Galton 1886). Using a one-group pre-post design, it is difficult to determine whether a change in participant outcomes is due to the intervention or to these confounding effects. The introduction of a comparison group can limit each of these threats. By measuring the changes in an equivalent comparison group from pre to post, which is also impacted by history, maturation, and regression

to the mean, the remaining differences between groups can be more confidently attributed to PHM program involvement.

Other common pre-post design elements that may limit study validity, even in a randomized setting, are the use of inadequate observation periods and few repeated measures to estimate the impact of PHM on cost. PHM evaluations frequently compare costs in the year prior to implementation to costs in the following year. Although this is an accepted method of evaluation, it may not adequately control for bias or other confounders. Employing a longitudinal design, with longer study periods and more frequent repeated cost measures, has been recommended as a better method for measuring changes over time in a natural environment (Linden, Adams, and Roberts 2003). Longitudinal data analysis can distinguish temporal changes within individuals from cross-sectional baseline group differences while accounting for the correlation between repeated observations. When the longitudinal relationship between PHM participation and cost is smaller than the cross-sectional relationship, the efficiency of the cost savings estimate can be improved by taking additional measurements on each person. Furthermore, increasing the within-subject variation in time with longer observation periods will increase the power, and thus the precision, of the savings estimate (Diggle et al. 2002). Longer follow-up periods can also help reduce any lingering effects of regression to the mean by continuing to measure group differences after the natural regression has occurred (Linden, Adams, and Roberts 2003).

One method of analysis that combines the benefits of longitudinal data with a quasi-experimental design is mixed model smoothing. This semiparametric longitudinal modeling approach uses both fixed and random effects to estimate risk-adjusted cost trends over time while accounting for autocorrelation. The flexibility of mixed model smoothing allows for the detection of group-specific cost trend fluctuations over time, while the smoothing component reduces the influence of outliers (Anderson and Jones 1995; Wang 1998; Wand 2003; Ngo and Wand 2004; Maringwa et al. 2008; Jacqmin-Gadda, Proust-Lima, and Amiéva 2010). In addition, inclusion of regular (e.g., quarterly) spline terms can aid in the measurement of natural changes due to seasonality and other history effects that impact both groups. Once these natural variations are removed, the remaining differences can be used to estimate the impact of PHM on costs.

In a quasi-experimental setting, identification of an adequate comparison group is also key to ensuring the internal validity of savings estimates. Without randomization, it is important to reduce and/or statistically control for preexisting group differences as much as possible to assess causality.

Selection bias is a common threat for quasi-experimental designs. When participants self-select their study group based on whether they choose to participate in PHM, selection bias is introduced. This is particularly concerning when those who choose to participate are inherently different from those who do not. Some of the methods for controlling selection bias include the use of a historical control group (Linden, Adams, and Roberts 2003; Dall et al. 2010), the selection of participations from other groups or sites that do not offer PHM services (Cousins and Liu 2003), and difference-in-difference models that use repeated observations to compare participants against themselves at earlier time points and then compare pre-post differences between groups (Nyman, Barleen, and Dowd 2009; Song et al. 2011).

The primary goal of the current research is to develop a consistent model for producing sound PHM cost savings estimates in a quasi-experimental setting. The framework should generate savings estimates that can account for natural temporal change, selection bias, and subject-specific regression to the mean effects and should be flexible enough to accommodate regular updates for the continued monitoring of PHM program effectiveness. A secondary goal is to assess the influence of sampling variability on the precision of the savings estimate to perform basic tests of statistical inference.

METHODS

The framework for estimating cost savings is developed using a commercial health plan population of employed individuals and their dependents enrolled between January 1, 2005 and June 30, 2009. Although a limited form of PHM services was available for health plan members since 2003, a more formal, structured PHM program was implemented in mid-2005. The new PHM program targeted members with at least one of the following chronic conditions: cardiovascular disease (i.e., hypertension, heart failure), asthma, diabetes, cancer, and chronic obstructive pulmonary disease (COPD). Two intervention intensity levels were available for members who opted in to the program: (1) a High Intensity level offering nurse case management, and (2) a Low Intensity level providing quarterly educational material regarding relevant health condition(s).

Participants

The current analysis includes adult (18+) health plan members enrolled during the study period with at least one of the conditions targeted by the PHM

program. Eligible members are assigned to one of four study groups based on their current and prior history of PHM participation. Two intervention groups consist of members currently enrolled in the PHM program each month stratified by intensity level. In addition, two comparison groups are available. Comparison Group 1 includes members who did not participate in PHM because they opted out, they could not be contacted, or their employer did not purchase PHM services. Comparison Group 2 consists of PHM enrollees during months when they are not enrolled.

Data

The data include health plan administrative data, PHM program records, and claims payment information spanning the period from January 2005 through June 2009. Data are aggregated at the month level for each member, where members could have up to 54 repeated observations based on the number of months they were enrolled in the health plan during the study period.

The predictors used to describe the cost trends include the following:

- Time: The primary term is a continuous measure of calendar *month* from 1 to 54, with 1 representing January 2005 and 54 representing June 2009. From this measure, quarterly knots are used to create linear spline terms spanning the entire study period. Each spline term is set to 0 until the knot, thereafter increasing by 1 each month through June 2009.
- Group: A set of dummy variables are used to identify the following study groups:
 - Never enrolled in PHM (Comparison Group 1)
 - Enrolled at PHM at some point during the study period but not enrolled in the current month (Comparison Group 2)
 - Currently enrolled in High Intensity PHM
 - Currently enrolled in Low Intensity PHM.

The group indicators are time dependent. Participants with a history of PHM enrollment can move between High Intensity, Low Intensity, and Comparison Group 2 depending on their enrollment status in the given month.

The outcome measure is *per member per month (PMPM) medical costs*, defined as the sum of all the medical claims incurred in a given month.¹ The covariates include age at program identification, gender, and health risk (as

measured by a 5-point ordinal measure of current resource use produced by the Johns Hopkins Adjusted Clinical Groups[®] Case-Mix System, where higher scores indicate higher resource utilization).

Analysis

The goal of the current framework is to elucidate the causal effect of PHM participation on costs. In theoretical terms, this is done by estimating the difference between observed (or factual) monthly costs for current PHM participants and their expected (or counterfactual) costs if they were not enrolled (Rubin 1974; Rosenbaum 1984; King and Zeng 2001). Although claims data can be used to determine the factual paid costs for members while enrolled in PHM, there are no cost records for these same members under an alternate scenario where they are not enrolled, and therefore the expected costs must be estimated. Ideally, in an experimental setting, the comparison group cost trend should represent the counterfactual costs for PHM participants if they had not enrolled. However, without the benefit of random assignment, the intervention and comparison groups may not be directly comparable due to baseline covariate differences that lead to biased cost trend estimates (Rosenbaum 1984). To assess the causal impact of PHM participation on costs in an observational setting, the current framework uses the cost trend for the comparison group as the expected cost trend estimate, although the baseline costs are adjusted to equal the baseline costs for PHM participants. This adjustment assists in isolating the differences in the cost trend over time that are related to program participation while parsing out the differences due to baseline covariate imbalances.

The first step in estimating PHM cost effectiveness is to estimate the PMPM cost trend for each study group. This is accomplished using mixed model smoothing. The model uses fixed effects for the covariates and a random intercept term to account for autocorrelation. A combination of fixed and random effects is employed to estimate group-specific cost trends; fixed effects for the linear time function and random effects for the quarterly spline terms of degree one. Both time functions are interacted with group to estimate group differences in costs over time.

The model is designed to estimate separate cost trends for each study group. The equation modeled to assess the relationship between PHM enrollment and PMPM costs is given as:

$$\begin{aligned}
\hat{E}(Y_{it}) = & \beta_{0i} + \beta_1(\text{Month})_t + \beta_2(\text{Not Currently Enrolled})_{it} \\
& + \beta_3(\text{Not Currently Enrolled} * \text{Month})_{it} + \beta_4(\text{Low Intensity})_{it} \\
& + \beta_5(\text{Low Intensity} * \text{Month})_{it} + \beta_6(\text{High Intensity})_{it} \\
& + \beta_7(\text{High Intensity} * \text{Month})_{it} + \beta_8(\text{Age})_i + \beta_9(\text{Female})_i \\
& + \beta_{10}(\text{Health Risk Score})_{it} \\
& + \left[\sum_{(k=1 \text{ to } K)} b_{1k}(\text{Month}_{it} - K_k)_+ \right]^* (\text{Never Enrolled})_{it} \\
& + \left[\sum_{(k=1 \text{ to } K)} b_{2k}(\text{Month}_{it} - K_k)_+ \right]^* (\text{Not Currently Enrolled})_{it} \\
& + \left[\sum_{(k=1 \text{ to } K)} b_{3k}(\text{Month}_{it} - K_k)_+ \right]^* (\text{Low Intensity})_{it} \\
& + \left[\sum_{(k=1 \text{ to } K)} b_{4k}(\text{Month}_{it} - K_k)_+ \right]^* (\text{High Intensity})_{it}
\end{aligned}$$

where Y_{it} represents the PMPM costs (Y) for person i ($i = 1 - N$) at month t ($t = 1-54$), $\beta_{0i} \sim N(0, \sigma^2_{\beta 0})$. The smoothing splines functions are estimated separately for each of the four study groups and are represented by the equation:

$$(\text{Month}_{it} - K_k)_+ = \begin{cases} 0, & \text{Month} \leq K_k \\ \text{Month} - K_k, & \text{Month} > K_k \end{cases}$$

where K_1, \dots, K_k are a set of quarterly spline knots ranging from 1 to 17 and $b_{gk} \sim N(0, \sigma^2[b_g])$ for $g = 1-4$ indicating the study group (Wand 2003).

Savings Calculation

After running the mixed model, the next step is to calculate the monthly mean “adjusted” (factual) PMPM costs for High and Low Intensity participants’ based on their model-based predicted values. Mean adjusted costs for month t , intensity j , is expressed by:

$$\hat{E}(Y | \text{Enrolled in PHM})_{tj} = \frac{\left[\sum_{i=1, i \in j}^n \hat{E}(Y_{it}) \right]}{n}$$

where j indicates either Low or High Intensity PHM participants and n represents the number of members enrolled each month by intensity level. To calculate expected (counterfactual) costs each month for these same participants if they had not been enrolled in PHM, the mean adjusted costs for Comparison Groups 1 and 2 are used, although there is a baseline shift so that costs in the first

implementation month are equal. Mean expected costs for month t , intensity j , comparison c , is calculated using the following formula::

$$\hat{E}(Y|\text{Not Enrolled})_{jc} = \frac{\left[\sum_{i=1, i \in c}^n \hat{E}(Y_{it}) \right]}{n} + \left[\hat{E}(Y)_{j,t=b} - \hat{E}(Y)_{c,t=b} \right]$$

where c represents Comparison Groups 1 or 2 and b represents the baseline month when the PHM program was first implemented (June to September 2005 or $t = 6-10$). Next, the difference D between adjusted and expected costs is calculated for each month t , intensity level j , and comparison group c , as represented by the equation:

$$\hat{E}(D)_{jc} = \hat{E}(Y|\text{Enrolled in PHM})_{jc} - \hat{E}(Y|\text{Not Enrolled})_{jc}$$

For each intensity level (j) and comparison group (c), the mean cost difference since the first implementation month is computed, weighted by the number of PHM members enrolled each month, as expressed by the following equation:

Cost Difference for Intensity j , Comparison c :

$$\hat{E}(D)_{jc} = \frac{\left[\sum_{t=b}^{54} \hat{E}(D)_{jc} * n_{jt} \right]}{\left[\sum_{t=b}^{54} n_{jc} \right]}$$

The cost difference for each intensity level is represented by the arithmetic average between the differences obtained using each of the comparison groups. The cost difference is then aggregated across intensity levels based on the average savings calculated for Low and High intensity, weighted by the proportion of members enrolled in each intensity level. Last, to reduce the sensitivity of the cost savings estimate to the starting month, the calculation of the total cost difference is replicated using five different starting months (June through October 2005 or $b = 6-10$), and the final savings estimate is the average cost difference from the five models weighted by the number of monthly observations used in the calculation. Using this framework, cost savings are realized when, on average, the adjusted costs for PHM participants are lower than expected (i.e., when the overall cost difference is less than 0).

Savings Estimate Precision

In a business setting, the cost savings estimate is usually the most meaningful measure of success, with less (if any) emphasis placed on statistical uncertainty

(Linden, Adams, and Roberts 2003). Nevertheless, the current analysis uses a modified bootstrapping approach to determine whether the difference between adjusted and expected costs is statistically significant. Rather than replicating the savings model 100 times from the full study population, as is done with traditional bootstrapping, this analysis samples participants, with replacement, so that the number in each group (i.e., enrollees and non-enrollees) matches the total number of members enrolled in PHM. This modification is made to reduce the processing time needed to perform bootstrapping (from approximately 2,200 to 96 hours). The results from the bootstrap samples are used to calculate the pooled standard error (SE) for adjusted and expected costs, although the SE for expected costs is adjusted for Comparison Group 1 to reflect that the bootstrap sample of comparison group members is smaller than the full sample (Rosner 2010). The adjusted pooled SE is estimated as follows:

$$\hat{SE}(D)_{jc} = \sqrt{[\hat{SE}(Y|\text{Enrolled in PHM})_j]^2 + [\hat{SE}(Y|\text{Enrolled in PHM})_{jc}]^2 * \frac{n_{\text{bootstrap}}}{n_{\text{full}}}}$$

where c includes Comparison Group 1 only. For Comparison Group 2, the traditional formula without adjustment is used to calculate the pooled SE. Lastly, the 95 percent confidence interval (CI) for the cost difference is calculated under the normality assumption using the adjusted pooled SE, represented by the equation:

$$95\% \text{ CI} = \hat{E}(D)_{jc} \pm 1.96\hat{SE}(D)_{jc}$$

A significant cost difference is reported if the 95 percent confidence interval does not include 0 (Briggs, Wonderling, and Mooney 1997).

The research study was approved by the Johns Hopkins University School of Medicine Institutional Review Board.

RESULTS

The current framework was developed using a total of 12,965 health plan members. The majority (10,804 or 83 percent) were never enrolled in the PHM program, primarily because they opted out or could not be reached, although a small portion (1 percent) did not participate due to ineligibility (i.e., their employer did not purchase PHM services). The intervention group

consisted of 2,161 members, with 51 percent in Low Intensity and 49 percent in High Intensity. These same participants served in Comparison Group 2 during months when they were not actively enrolled (77 percent of 96,558 total member months). A description of the intervention and Comparison Group 1 participants is included in Table 1.

Savings Rate

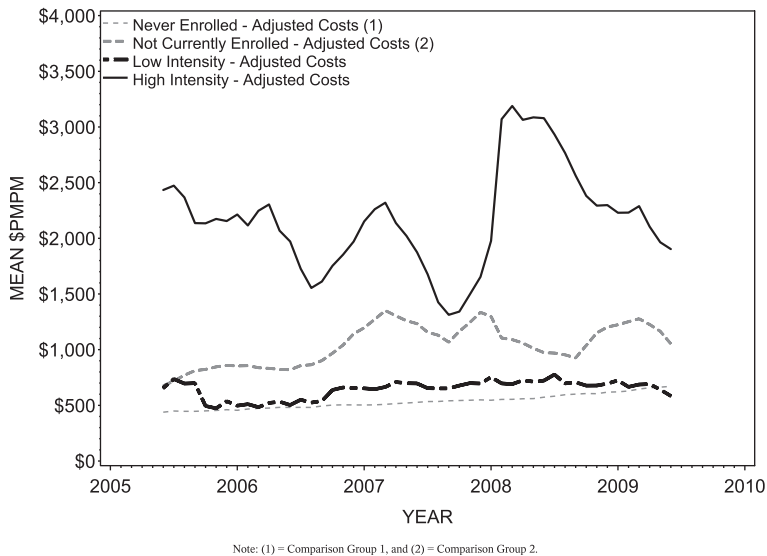
Comparisons of actual costs, expected costs, and cost savings are represented graphically in Figures 1–4, with the cost savings estimates listed in Table 2. In June 2005, the first of five baseline months, mean costs for Comparison Group 1, Comparison Group 2, Low Intensity, and High Intensity members are \$438.49, \$675.29, \$655.62, and \$2,433.72, respectively. The differences suggest that on average, High Intensity members start off with higher costs than either of the comparison groups, particularly for non-enrollees (Comparison Group 1). Comparison group baseline costs are more comparable for Low Intensity members (Figure 1). (See Appendix A for illustrations of model fit using actual and adjusted cost trends by group.)

Figure 2 illustrates the average monthly medical costs for all High Intensity PHM participants from June 2005 through June 2009, adjusted for age, gender, and health risk. Also shown are the expected cost trends for these same members if they were not enrolled, as estimated by each comparison

Table 1: Participant Demographics by Study Group

	<i>Intervention</i>		<i>Comparison Group 1</i>
	<i>High Intensity</i>	<i>Low Intensity</i>	
Members	1,069	1,092	10,804
Mean age at baseline (in years)	47.4	47.4	43.1
Percent female	70	67	63
Mean number of months in health plan	44.6	44.8	40.6
Prevalence of chronic conditions (%)			
Cardiovascular disease	90	86	79
Hypertension	81	73	62
Heart failure	8	3	2
Diabetes	64	30	24
Asthma	28	24	24
COPD	11	5	4
Coronary artery disease	25	13	10
Kidney disease	13	4	3
Cancer	1	1	5
Multiple (2+) chronic conditions (%)	90	81	71

Figure 1: Adjusted Per Member Per Month Cost Trends by Group

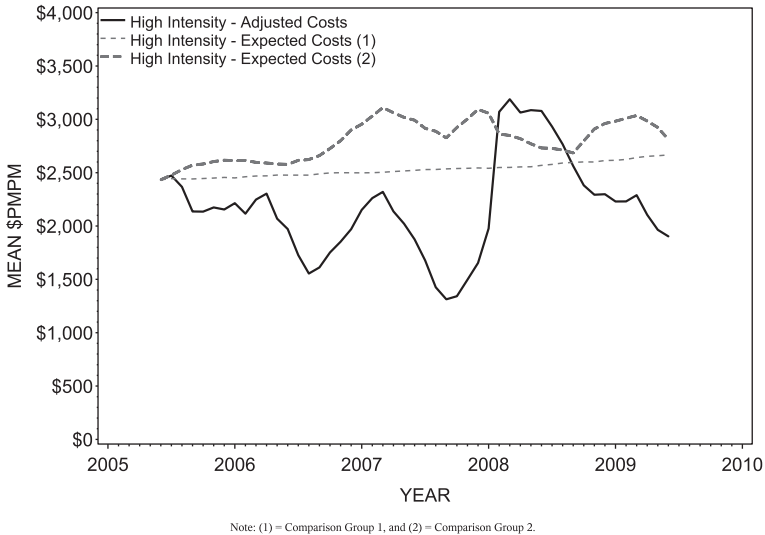


group. Although the graph uses June 2005 for illustrative purposes, the final savings rate is calculated using five different starting months (June through October; see Appendix B for a full list of savings estimates by starting month). In general, adjusted costs are lower than expected given the cost trends for the comparison groups, particularly when considering the expected cost trend based on members more likely to enroll (Comparison Group 2). The average cost difference obtained using each of the expected cost trends is $-\$277.93$ ($p \geq .05$), indicating that participants in the High Intensity intervention incur monthly costs that are $\$277.93$ lower than expected without the PHM program, although the difference is not statistically significant.

The next figure (Figure 3) replicates the previous illustration with the adjusted and expected pmpm medical costs for Low Intensity members. Similar to the findings for High Intensity, the cost differences for Low Intensity vary based on the comparison group used to estimate the savings. On average, pmpm costs for Low Intensity members are $\$247.39$ lower than would have been expected without PHM involvement, and this difference is statistically significant ($p < .05$).

The monthly differences between actual and expected costs are depicted by comparison group and intensity level in Figure 4. The cumulative cost difference across intensity levels is $-\$145.63$ ($p \geq .05$) based on Comparison Group 1 expected costs and $-\$380.30$ ($p \geq .05$) based on Comparison Group 2.

Figure 2: Adjusted and Expected Per Member Per Month Costs for High Intensity Members Since the Start of the Population Health Management Program



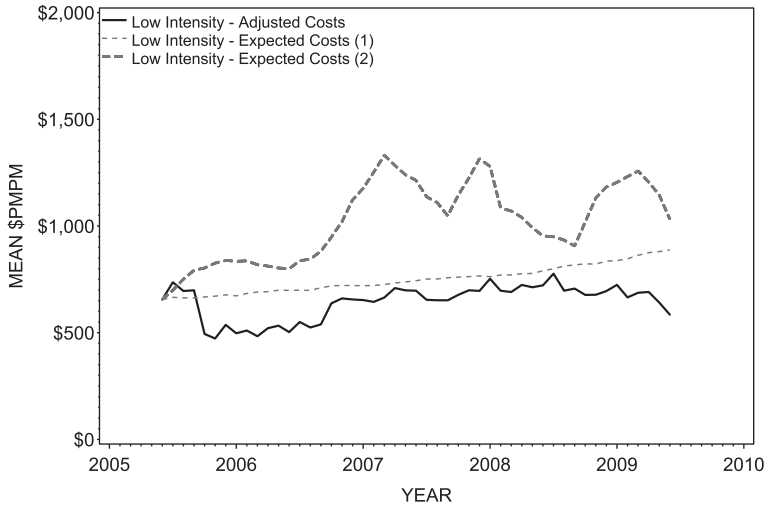
The weighted average cost difference across intensity level and comparison group is $-\$262.96$ per participant per month ($p \geq .05$). This indicates that, on average, PHM participants cost $\$262.96$ less than they would have without PHM program services, although this difference is not statistically significant.

DISCUSSION

The current study provides a sound, consistent method for calculating PHM cost savings estimates using observational data. This method reduces the threats to internal validity that have plagued much of the prior research on PHM cost effectiveness.

First, rather than using a simple pre-post design with only two repeated measures, the current framework involves a longitudinal data analysis of monthly observations spanning over 4 years. Longer study periods help to improve model fit (Linden, Adams, and Roberts 2003), and studies with longer follow-up periods allow for the detection of changes to PHM program cost effectiveness over longer implementation periods. However, even when

Figure 3: Adjusted and Expected Per Member Per Month Costs for Low Intensity Members Since the Start of the Population Health Management Program



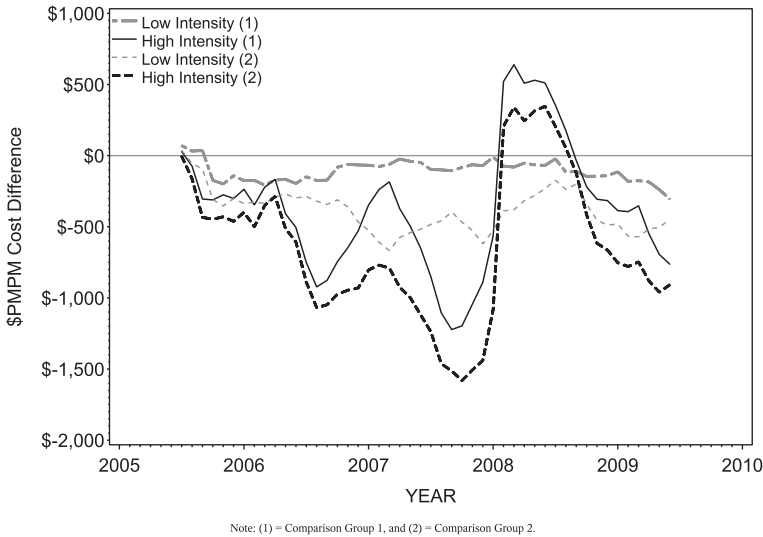
Note: (1) = Comparison Group 1, and (2) = Comparison Group 2.

compared to other longitudinal designs with multiple yearly outcome measures, the current framework's use of monthly outcomes increases the number of observations, which further enhances the model's ability to characterize cost trends over time. In addition to improving model fit, monthly observations enable the detection of smaller variations in cost trends that would likely have been lost when averaging across year-long observation periods.

The validity of the current framework is also enhanced by the quasi-experimental design. The identification of a similar comparison group limits the threats of history, maturation, and regression to the mean that are common to observational studies lacking such a comparison.

Without randomization, however, it is difficult to tell with certainty whether cost differences are a result of the program or of preexisting differences between those who choose to participate and those who refuse (e.g., selection bias). In the current study, comparisons of the baseline costs suggest that the groups were not equivalent at baseline. The selection bias threat is partially limited by the inclusion of a small number of members in Comparison Group 1 who did not enroll because their employer did not purchase PHM services. The uncertainty caused by selection bias is also limited by

Figure 4: Adjusted and Expected Per Member Per Month Cost Differences by Intensity Level and Comparison Group



focusing on differences in trends, rather than observed mean differences. The current model accounts for baseline differences by shifting the comparison group cost trend to meet the baseline costs for intervention members. Because the baseline mean costs for Comparison Group 1 members are lower than those demonstrated for intervention members, it is likely that using the cost trend for nonparticipants underestimates the expected trend for the intervention members. Thus, it provides a more conservative estimate of the savings.

To further overcome the uncertainty caused by selection bias, a second comparison group is identified. Comparison Group 2 reduces baseline group differences in morbidity, risk, and utilization as it is made up of the exact same members as the intervention group during months when they are not enrolled in PHM. The cost trend for this comparison group is often steeper than that found for Comparison Group 1, which supports the notion that the first comparison group underestimates expected costs. As Comparison Group 2 is more similar to intervention members, it could be argued that the cost difference obtained using this comparison group is a better estimate of the true savings. Instead, the current framework takes a more conservative approach to cost savings estimation by taking an average of the cost differences calculated using both comparison groups.

Table 2: Adjusted and Expected Cost Differences in Dollars

<i>Cost Difference by Intensity Level (\$)</i>	<i>Comparison Group 1 Difference (95% CI)</i>	<i>Comparison Group 2 Difference (95% CI)</i>	<i>Total Difference (95% CI)</i>
Low intensity	−120.83 (−229.41, −12.25)*	−373.96 (−593.68, −154.24)*	−247.39 (−411.61, −83.18)*
High intensity	−169.48 (−697.69, 358.73)	−386.39 (−1,340.68, 567.91)	−277.93 (−1,025.20, 469.34)
Total	−145.63 (−416.73, 125.47)	−380.30 (−878.26, 117.67)	−262.96 (−647.52, 121.60)

Note. Negative cost differences indicate Population Health Management program savings. Total savings rates for each intensity level are computed based on the mean savings rate across comparison groups. Savings are aggregated across intensity levels based on the average savings rate for low and high intensity weighted by the proportion of intervention members enrolled in each intensity level.

*Statistically significant cost differences.

The current framework’s operationalization of time in calendar months, rather than a more pre-post approach, further enhances its internal validity. With quasi-experimental pre-post designs, regression to the mean can still limit study findings, particularly when baseline group differences exist. If pre-intervention measures indicate that costs are typically higher for intervention members than for the comparison group, then regression to the mean suggests that regardless of PHM involvement, the decline in costs with each subsequent measurement following implementation is likely to be more extreme for the intervention group because their costs started further from the mean. A probable consequence of this method is overinflated savings estimates that may be partially or fully explained by regression to the mean. The current study limits this threat by modeling time in calendar months instead of months since PHM enrollment. By doing so, intervention group costs are characterized by the average costs across members with different lengths of program enrollment, from newly enrolled members to those who have been enrolled for years, and thus the threat of regression to the mean is significantly mitigated.

Another important enhancement introduced by this method is its flexibility to accommodate regular replications for the continued monitoring of PHM program effectiveness. With the utilization of monthly observations, the savings rate can be updated each month as more claims data become available. The savings rate can also be used to estimate the ROI, which is another important funding consideration.

Although the current framework includes multiple enhancements to preexisting cost savings methodologies, additional improvements could be made to strengthen both internal and external validity. First, when identifying

comparison group members, propensity scores could be incorporated into the sampling process to further enhance group comparability at baseline. Second, the use of alternative smoothing techniques that place a greater penalty on deviations from the mean could reduce variability. Improving the precision of the estimate could help to increase the model's ability to detect significant cost differences. Last, the current study would need to be replicated using populations other than employed members of a commercial health plan to ensure the generalizability of the framework to older or unemployed populations (e.g., Medicare, Medicaid).

By utilizing this framework for estimating PHM savings, national, state, and private payers who hold the risk for health care expenditures will have a confident basis on which to make investment decisions. Ideally, through the provision of effective, cost-efficient PHM services to those in need, the health of multimorbid populations can be improved and the rising trend in health care costs can be contained or even reduced.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was conducted as part of an internal health plan evaluation of Population Health Management cost effectiveness. The authors would like to thank Paula Norman for compiling the research data set from existing records and Kyle Marmen for providing guidance related to health plan enrollment and benefit structures.

Disclosures: None.

Disclaimers: None.

NOTE

1. The outcome is limited to medical costs due to the lack of pharmacy claims data for the entire study duration. Extensive analyses were conducted to compare the total and medical-cost-only models (where pharmacy data were available), and the pmpm savings rates only differed by approximately \$10. Thus, the medical cost model is considered a close approximation of the total cost model.

REFERENCES

- Anderson, S. J., and R. H. Jones. 1995. "Smoothing Splines for Longitudinal Data." *Statistics in Medicine* 14 (11): 1235–48.
- Briggs, A. H., D. E. Wonderling, and C. Z. Mooney. 1997. "Pulling Cost-Effectiveness Analysis up by Its Bootstraps: A Non-Parametric Approach to Confidence Interval Estimation." *Health Economics* 6 (4): 327–40.
- Chapman, L. S., and K. R. Pelletier. 2004. "Population Health Management as a Strategy for Creation of Optimal Healing Environments in Worksite and Corporate Settings." *The Journal of Alternative and Complementary Medicine* 10 (S1): 127–40.
- Cloutier, M. M., S. D. Grosse, D. B. Wakefield, T. A. Nurmagambetov, and C. M. Brown. 2009. "The Economic Impact of an Urban Asthma Management Program." *The American Journal of Managed Care* 15 (6): 345–51.
- Congressional Budget Office. 2004. An Analysis of the Literature on Disease Management Programs. [accessed on October 5, 2010]. Available at <http://www.cbo.gov>
- Cook, T. D., and D. T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Cousins, M. S., and Y. Liu. 2003. "Cost Savings for a Preferred Provider Organization Population with Multi-Condition Disease Management: Evaluating Program Impact Using Predictive Modeling with a Control Group." *Disease Management* 6 (4): 207–17.
- Dall, T. M., R. C. Askarinam Wagner, Y. Zhang, W. Yang, D. R. Arday, and C. J. Gantt. 2010. "Outcomes and Lessons Learned from Evaluating TRICARE's Disease Management Programs." *The American Journal of Managed Care* 16 (6): 438–46.
- Diggle, P. J., P. Heagerty, K. Liang, and S. L. Zeger. 2002. *Analysis of Longitudinal Data*, 2nd Edition. New York: Oxford University Press.
- Fetterolf, D., D. Wennberg, and A. Devries. 2004. "Estimating the Return on Investment in Disease Management Programs Using a Pre-Post Analysis." *Disease Management* 7 (1): 5–23.
- Fitzner, K., J. Sidorov, D. Fetterolf, D. Wennberg, E. Eisenberg, M. Cousins, J. Hoffman, J. Houghton, W. Charlton, D. Krause, A. Woolf, K. McDonough, W. Todd, K. Fox, D. Plocher, I. Juster, M. Stiefel, V. Villagra, and I. Duncan. 2004. "Principles for Assessing Disease Management Outcomes." *Disease Management* 7 (3): 191–201.
- Galbreath, A. D., R. A. Krasuski, B. Smith, K. C. Stajduhar, M. D. Kwan, R. Ellis, et al. 2004. "Long-Term Healthcare and Cost Outcomes of Disease Management in a Large, Randomized, Community-Based Population with Heart Failure." *Circulation* 110 (23): 3518–26.
- Galton, F. 1886. "Regression Towards Mediocrity in Hereditary Stature." *The Journal of the Anthropological Institute of Great Britain and Ireland* 15: 246–63.
- Gary, T. L., M. Batts-Turner, H. C. Yeh, F. Hill-Briggs, L. R. Bone, N. Y. Wang, et al. 2009. "The Effects of a Nurse Case Manager and A Community Health Worker Team on Diabetic Control, Emergency Department Visits,

- and Hospitalizations among Urban African Americans with Type 2 Diabetes Mellitus: A Randomized Controlled Trial." *Archives of Internal Medicine* 169 (19): 1788–94.
- Hibbard, J. H., J. Greene, and M. Tusler. 2009. "Improving the Outcomes of Disease Management by Tailoring Care to the Patient's Level of Activation." *The American Journal of Managed Care* 15 (6): 353–60.
- Hoffman, C., D. Rice, and H. Y. Sung. 1996. "Persons with Chronic Conditions: Their Prevalence and Costs." *Journal of the American Medical Association* 276 (18): 1473–9.
- Jacqmin-Gadda, H., C. Proust-Lima, and H. Amiéva. 2010. "Semi-Parametric Latent Process Model for Longitudinal Ordinal Data: Application to Cognitive Decline." *Statistics in Medicine* 29 (26): 2723–31.
- Johnson, A. 2003. *Disease Management: The Programs and the Promise*. Milliman Research Report. [accessed on August 16, 2012]. Available at: <http://publications.milliman.com/research/health-rr/archive/pdfs/Disease-Mangement-Programs-Promise-RR05-01-03.pdf>.
- Juster, I. A., S. N. Rosenberg, D. Senapati, and M. R. Shah. 2009. "'Dial-an-ROI?' Changing Basic Variables Impacts Cost Trends in Single-Population Pre-Post ('DMAA type') Savings Analysis." *Population Health Management* 12 (1): 17–24.
- Kindig, D., and G. Stoddart. 2003. "What Is Population Health?" *American Journal of Public Health* 93 (3): 380–3.
- King, G., and L. Zeng. 2001. *How Factual Is Your Counterfactual?* Cambridge, MA: Harvard Burden of Disease Unit.
- Linden, A. L. 2006. "What Will It Take for Disease Management to Demonstrate a Return on Investment? New Perspectives on an Old Theme." *The American Journal of Managed Care* 12 (4): 217–22.
- Linden, A., J. L. Adams, and N. Roberts. 2003. "Evaluating Disease Management Program Effectiveness: An Introduction to Time-Series Analysis." *Disease Management* 6 (4): 243–55.
- Maciejewski, M. L., S. Y. Chen, and D. H. Au. 2009. "Adult Asthma Disease Management: An Analysis of Studies, Approaches, Outcomes, and Methods." *Respiratory Care* 54 (7): 878–86.
- Maringwa, J. T., H. Geys, Z. Shkedy, C. Faes, G. Molenberghs, M. Aerts, et al. 2008. "Application of Semiparametric Mixed Models and Simultaneous Confidence Bands in a Cardiovascular Safety Experiment with Longitudinal Data." *Journal of Biopharmaceutical Statistics* 18 (6): 1043–62.
- McCall, N., and J. Cromwell. 2011. "Results of the Medicare Health Support Disease-Management Pilot Program." *New England Journal of Medicine* 365 (18): 1704–12.
- Ngo, L., and M. P. Wand. 2004. "Smoothing with Mixed Model Software." *Journal of Statistical Software* 9 (1): 1–54.
- Nyman, J. A., N. A. Barleen, and B. E. Dowd. 2009. "A Return-on-Investment Analysis of the Health Promotion Program at the University of Minnesota." *Journal of Occupational and Environmental Medicine/American College of Occupational and Environmental Medicine* 51 (1): 54–65.

- Peikes, D., A. Chen, J. Schore, and R. Brown. 2009. "Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures among Medicare Beneficiaries: 15 Randomized Trials." *Journal of the American Medical Association* 301 (6): 603–18.
- Rosenbaum, P. R. 1984. "From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment." *Journal of the American Statistical Association* 79 (385): 41–8.
- Rosner, B. 2010. *Fundamentals of Biostatistics*, 7th Edition. Boston, MA: Cengage Learning, Inc.
- Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.
- Selby, J. V., D. Scanlon, J. E. Lafata, V. Villagra, J. Beich, and P. R. Salber. 2003. "Determining the Value of Disease Management Programs." *Joint Commission Journal on Quality and Safety* 29 (9): 491–9.
- Seto, E. 2008. "Cost Comparison between Telemonitoring and Usual Care of Heart Failure: A Systematic Review." *Telemedicine Journal and e-Health: The Official Journal of the American Telemedicine Association* 14 (7): 679–86.
- Sidorov, J., R. Shull, J. Tomcavage, S. Girolami, N. Lawton, and R. Harris. 2002. "Does Diabetes Disease Management Save Money and Improve Outcomes? A Report of Simultaneous Short-Term Savings and Quality Improvement Associated with a Health Maintenance Organization-Sponsored Disease Management Program among Patients Fulfilling Health Employer Data and Information Set Criteria." *Diabetes Care* 25 (4): 684–9.
- Song, Z., D. G. Safran, B. E. Landon, Y. He, R. P. Ellis, R. E. Mechanic, et al. 2011. "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract." *The New England Journal of Medicine* 365 (10): 909–18.
- Sylvia, M. L., M. Griswold, L. Dunbar, C. M. Boyd, M. Park, and C. Boulton. 2008. "Guided Care: Cost and Utilization Outcomes in a Pilot Study." *Disease Management* 11 (1): 29–36.
- Turner, D. A., S. Paul, M. A. Stone, A. Juarez-Garcia, I. Squire, and K. Khunti. 2008. "Cost-Effectiveness of a Disease Management Programme for Secondary Prevention of Coronary Heart Disease and Heart Failure in Primary Care." *Heart (British Cardiac Society)* 94 (12): 1601–6.
- Villagra, V. G., and T. Ahmed. 2004. "Effectiveness of a Disease Management Program for Patients with Diabetes." *Health Affairs (Project Hope)* 23 (4): 255–66.
- Wagner, E. H. 1998. "Chronic Disease Management: What Will It Take to Improve Care for Chronic Illness?" *Effective Clinical Practice* 1 (1): 2–4.
- Wand, M. P. 2003. "Smoothing and Mixed Models." *Computational Statistics* 18 (2): 223–50.
- Wang, Y. 1998. "Smoothing Spline Models with Correlated Random Errors." *Journal of the American Statistical Association* 93 (441): 341–8.

- Wheatley, B. 2002. "Disease Management: Findings from Leading State Programs." *State Coverage Initiatives Issue Brief: A National Initiative of the Robert Wood Johnson Foundation* 3 (3): 1–6.
- Wilson, T. W., J. Gruen, T. William, D. Fetterolf, P. Minalkumar, R. G. Popiel, et al. 2004. "Assessing Return on Investment of Defined-Population Disease Management Interventions." *Joint Commission Journal on Quality and Safety* 30 (11): 614–21.
- Wolff, J. L., B. Starfield, and G. Anderson. 2002. "Prevalence, Expenditures, and Complications of Multiple Chronic Conditions in the Elderly." *Archives of Internal Medicine* 162 (20): 2269–76.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Model Fit.

Appendix SA3: Sensitivity to the Starting Month.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.